



dr hab. inż. Piotr A. Kowalski, prof. AGH
Katedra Informatyki Stosowanej i Fizyki Komputerowej,
Wydział Fizyki i Informatyki Stosowanej,
Akademia Górniczo-Hutnicza w Krakowie,
al. Mickiewicza 30, 30-059 Kraków
email: pkowal@agh.edu.pl

oraz



Centrum Informatycznych Metod Analizy Danych
Instytut Badań Systemowych
Polskiej Akademii Nauk

Kraków, 15.07.2026

RECENZJA

**Recenzja rozprawy doktorskiej mgr. inż. Michała Daniluka
pt. „Multimodal recommendation system for e-commerces”**

1. Charakterystyka rozprawy

Przedmiotem niniejszej recenzji jest rozprawa doktorska mgr. inż. Michała Daniluka pt. „*Multimodal recommendation system for e-commerce*”, przygotowana na Politechnice Warszawskiej w dyscyplinie informatyka techniczna i telekomunikacja, pod kierunkiem prof. ucz. dr. hab. inż. Grzegorza Pastuszaka. Praca została napisana w języku angielskim, przy czym zawiera również streszczenie w języku polskim. Jej tematyka dotyczy systemów rekomendacyjnych w e-commerce, ze szczególnym uwzględnieniem reprezentacji wielomodalnych, estymacji gęstości, modeli fundacyjnych dla danych behawioralnych oraz efektywnego przetwarzania dużych wolumenów danych użytkowników.

Rozprawa ma przejrzystą, logiczną strukturę i składa się z ośmiu rozdziałów. Rozdział pierwszy wprowadza problem badawczy, formułuje cele, hipotezy oraz przedstawia wkład autora. Rozdział drugi przedstawia podstawy systemów rekomendacyjnych, w tym filtrowanie kolaboracyjne, metody oparte na treści, problem zimnego startu, rzadkość danych, skalowalność oraz zagadnienia ewaluacji. Rozdział trzeci omawia stan badań, obejmując modele sekwencyjne, grafowe sieci neuronowe, rekomendacje wielomodalne oraz próby budowy modeli fundacyjnych dla rekomendacji. Rozdział czwarty poświęcono reprezentacjom grafowym i metodzie Cleora, która stanowi ważny element tła technicznego dla dalszych części pracy. Rozdział piąty przedstawia metodę EMDE, czyli Efficient Manifold Density Estimator. Rozdział szósty zawiera opis najważniejszego wkładu rozprawy, tj. BaseModel - samonadzorowany model dla danych behawioralnych. Rozdział siódmy przedstawia walidację metod w konkursach międzynarodowych, natomiast rozdział ósmy zawiera podsumowanie, ocenę hipotez,

ograniczenia oraz kierunki dalszych badań. Bibliografia rozpoczyna się na stronie 153 i obejmuje 90 pozycji, w tym prace z zakresu systemów rekomendacyjnych, modeli grafowych, uczenia samonadzorowanego, reprezentacji multimodalnych oraz zastosowań przemysłowych.

2. Teza i cel pracy

Rozprawa nie formułuje jednej tezy w postaci osobnego, krótkiego zdania, jednak jej zasadniczą tezę można odtworzyć bardzo wyraźnie. Autor przyjmuje strategię dystrybucyjną, zgodnie z którą historia interakcji użytkownika stanowi próbkę z ukrytego rozkładu preferencji w przestrzeni osadzeń produktów, a nie jedynie sekwencję zdarzeń ani punktową reprezentację otrzymaną przez uśrednianie wektorów produktów. W takim ujęciu istotne nie jest wyłącznie pytanie, w jakiej kolejności użytkownik klikał lub kupował produkty, ale raczej: jakie obszary przestrzeni produktów odwiedzał, z jaką intensywnością i jak zmieniały się jego preferencje w czasie. Tezę tę autor rozwija poprzez pięć hipotez badawczych dotyczących przewagi reprezentacji gęstościowych nad punktowymi, integracji multimodalnej poprzez konkatenację szkiców, ograniczenia potrzeby stosowania złożonych architektur, transferu dzięki pretrenowaniu samonadzorowanemu oraz korzyści w scenariuszach zimnego startu.

3. Streszczenie rozprawy

Streszczając zasadniczy tok rozprawy, należy podkreślić, że autor zaczyna od krytyki dominującego w ostatnich latach paradygmatu sekwencyjnego. Wskazuje, że w języku naturalnym kolejność słów ma zasadnicze znaczenie semantyczne, natomiast w wielu scenariuszach zakupowych kolejność kliknięć lub zakupów może wynikać z układu strony, promocji, dostępności produktu albo przypadku. Z tego powodu modelowanie użytkownika jako sekwencji tokenów może wprowadzać do procesu uczenia strukturę, która nie zawsze odzwierciedla rzeczywistą dynamikę preferencji. Alternatywą proponowaną w pracy jest reprezentacja dystrybucyjna, a historia zachowań użytkownika jest traktowana jako próbka z rozkładu preferencji w przestrzeni osadzeń produktów.

Pierwszym ważnym składnikiem tego podejścia jest EMDE, metoda łącząca idee haszowania wrażliwego na lokalność oraz szkiców zliczających, uzupełniona o podział zależny od gęstości danych. Dzięki temu historie użytkowników mogą być przedstawiane jako szkice o stałym rozmiarze, zachowujące wielomodalną strukturę preferencji. Drugim, centralnym osiągnięciem pracy jest BaseModel -model samonadzorowany, który przewiduje przyszłe rozkłady zachowań na podstawie rozkładów przeszłych. Autor pokazuje, że jeden wstępnie wytrenowany rdzeń może obsługiwać zadania rekomendacji, klasyfikacji i regresji z użyciem lekkich głowic zadaniowych.

4. Najważniejsze osiągnięcia rozprawy

Do najważniejszych osiągnięć rozprawy zaliczam przede wszystkim zaproponowanie spójnego sposobu reprezentacji historii behawioralnych jako rozkładów, a nie pojedynczych punktów w przestrzeni cech. Jest to koncepcyjnie przekonujące, dobrze umotywowane i praktycznie użyteczne. Drugim istotnym osiągnięciem jest rozwinięcie metody EMDE w kierunku wielomodalnej reprezentacji danych, w której szkice pochodzące z różnych modalności mogą być łączone poprzez prostą konkatenację. Trzecim osiągnięciem jest opracowanie BaseModel jako samonadzorowanego modelu dla danych behawioralnych. Czwartym - szeroka walidacja empiryczna, obejmująca benchmarki akademickie, zadania RelBench, porównania z modelami TIGER i HSTU oraz analizy zimnego startu. Piątym - wyniki konkursowe: drugie miejsce w Booking.com WSDM Data Challenge, trzecie miejsce w KDD Cup oraz drugie miejsce w ACM RecSys Challenge. Szóstym - rzeczywiste wdrożenia przemysłowe, w tym przetwarzanie 42 miliardów zdarzeń miesięcznie dla ponad 200 organizacji oraz raportowane wzrosty metryk biznesowych w zastosowaniach e-commerce, finansowych i telekomunikacyjnych.

Na szczególne uznanie zasługuje sposób, w jaki autor łączy refleksję naukową z praktycznym doświadczeniem wdrożeniowym. Rozprawa nie ogranicza się do testów na standardowych zbiorach danych, lecz pokazuje także, że proponowane rozwiązania mogą funkcjonować pod presją wymagań produkcyjnych: niskiej latencji, dużej skali, brakujących danych, konieczności częstego odświeżania modeli oraz ograniczeń regulacyjnych. Ten aspekt nadaje pracy bardzo wyraźny charakter aplikacyjny, ale bez utraty waloru naukowego.

5. Ocena rozprawy

a. Uwagi krytyczno-polemiczne:

Moja ocena rozprawy jest jednoznacznie pozytywna. Praca jest ambitna, wielowątkowa, dobrze osadzona w aktualnym stanie badań i zawiera zarówno wkład koncepcyjny, jak i metodyczny oraz eksperymentalny. Poniższe uwagi traktuję jako punkty do dyskusji, które mogą być inspirujące podczas obrony lub w dalszych pracach Autora.

1. Autor przekonująco pokazuje, że odpowiednia reprezentacja może zmniejszać potrzebę stosowania złożonych architektur. Miejscami jednak narracja przyjmuje bardzo silną postać, w której reprezentacja ma dominować nad architekturą. Warto byłoby jeszcze wyraźniej doprecyzować, w jakich klasach problemów ta teza jest najmocniejsza, a gdzie modele sekwencyjne, grafowe lub oparte na mechanizmie uwagi mogą nadal być naturalnym wyborem.
2. Pojęcie „foundation model” jest używane ostrożnie, ale nadal może budzić kontrowersje. Autor sam zaznacza, że BaseModel nie jest modelem fundacyjnym w takim sensie, w jakim

termin ten funkcjonuje w odniesieniu do dużych modeli językowych. Jest to uczciwe i potrzebne zastrzeżenie. Jednocześnie warto byłoby jeszcze bardziej oddzielić model ten, trenowany osobno dla danego zbioru lub domeny, od modelu fundacyjnego zdolnego do szerokiego transferu między domenami.

3. Walidacja produkcyjna jest bardzo cenna, ale z natury rzeczy ograniczona przez poufność. Wyniki wdrożeniowe są jednym z najmocniejszych elementów pracy. Jednocześnie czytelnik nie zawsze ma dostęp do szczegółów eksperymentów produkcyjnych: wielkości grup, poziomów istotności statystycznej, wariancji efektów, długości testów, kryteriów wykluczeń czy pełnej specyfikacji grup kontrolnych. Zatem warto byłoby rozważyć syntetyczny, ale bardziej formalny opis protokołu wdrożeniowego.
4. Porównania z bardzo dużymi modelami przemysłowymi wymagają dodatkowego komentarza metodologicznego. Wyniki względem HSTU i TIGER są imponujące, jednak porównania z modelami o bardzo różnej skali, architekturze i sposobie raportowania zawsze rodzą pytania o identyczność protokołów, dobór hiperparametrów, implementację bazową oraz sprawiedliwość warunków obliczeniowych. Autor omawia te kwestie, ale można byłoby je jeszcze bardziej uwypuklić w osobnym podsumowaniu metodologicznym.
5. Teoretyczne własności EMDE mogłyby zostać rozwinięte bardziej szczegółowo. Intuicja stojąca za szkicami gęstościowymi jest bardzo przekonująca, podobnie jak empiryczna walidacja. Ale warto byłoby rozwinąć formalną analizę własności aproksymacyjnych EMDE, zwłaszcza wpływu liczby niezależnych partycji (**number of independent partitionings, sketch depth, N**), liczby kubeków (**number of buckets per partitioning, sketch width, B**), liczby hiperplanów (**number of hyperplanes per partitioning, K**), kolizji haszowania (**hash collisions**) oraz losowości hiperplanów (**random hyperplanes, random directions**) na stabilność i błąd reprezentacji.
6. Integracja multimodalna przez konkatenację jest elegancka, ale warto byłoby dodać więcej analiz awarii. Rozprawa pokazuje, że prosta konkatenacja szkiców działa dobrze. Dyskusyjnie można jednak zapytać, co dzieje się w sytuacjach silnie nie zrównoważonych modalności: gdy jedna modalność jest bardzo zaszumiona, inna niekompletna, a trzecia dominuje informacyjnie. Bardziej szczegółowa analiza przypadków brakujących modalności, degradacji jakości obrazów, ubogich opisów produktów lub silnego szumu metadanych byłaby wartościowym uzupełnieniem.
7. Wątek etyczny, prywatnościowy i związany z pętlami sprzężenia zwrotnego mógłby zostać bardziej zaakcentowany. Systemy rekomendacyjne wpływają na zachowania użytkowników i mogą wzmacniać popularność wybranych produktów, ograniczać różnorodność rekomendacji lub utrzymywać istniejące nierówności widoczności produktów. Autor wspomina o problemach

ewaluacji i produkcyjnego wdrożenia, ale w pracy o tak wyraźnym zastosowaniu przemysłowym warto byłoby szerzej omówić kwestie prywatności, fairness, bias, efekt bańki rekomendacyjnej oraz długoterminowe skutki optymalizacji metryk biznesowych.

8. Rozdział dotyczący Cleora jest konieczny, ale granica między wkładem własnym a tłem mogłaby być jeszcze wyraźniej zaznaczona w strukturze pracy. Autor uczciwie wskazuje, że Cleora została opracowana przez współpracowników i stanowi tło dla dalszych metod. Ponieważ jednak rozdział ten zajmuje istotne miejsce w rozprawie, czytelnikowi pomocne byłoby jeszcze mocniejsze wizualne i redakcyjne odróżnienie elementów własnych od elementów kontekstowych, np. poprzez wyraźną tabelę „wkład własny / elementy wykorzystane”.

W szczególności z racji na obowiązki recenzenta, muszę w tym miejscu wskazać jeszcze kilka uwag związanych z pojedynczymi niedociągnięciami stylistycznymi. Doktorant również nie ustrzegł się nielicznych mankamentów technicznych, takich jak błędy interpunkcyjne. W pracy występują również nieliczne usterki edytorskie takie jak np.

- **Strona 22 -brak separatora po informacji o punktach MEiN.**

Fragment „MEiN: 200 points Foundational work...” powinien zawierać kropkę lub średnik po „points”, np. „MEiN: 200 points. Foundational work...”.

- **Strona 22 -publikacja „Under review” umieszczona wśród publikacji konferencyjnych z punktacją.**

Pozycja „BaseModel: A Foundation Model for Behavioral Data. Under review at ACM RecSys 2026. MEiN: 140 points” może być myląca. Praca w recenzji nie powinna być prezentowana na równi z przyjętymi publikacjami recenzowanymi, a ewentualna punktacja powinna zostać opisana jako potencjalna lub usunięta do czasu akceptacji.

Powyższe uwagi mają charakter porządkujący i nie podważają wartości pracy.

6. Podsumowanie

Przedstawiona rozprawa doktorska jest pracą bardzo dojrzałą, ambitną i wyraźnie osadzoną zarówno w aktualnym nurcie badań nad systemami rekomendacyjnymi, jak i w realnych wymaganiach systemów produkcyjnych. Jej największą wartością jest konsekwentne przesunięcie akcentu z coraz bardziej złożonych architektur na jakość reprezentacji danych behawioralnych. Autor pokazuje, że odpowiednio zaprojektowana reprezentacja dystrybucyjna może nie tylko zachować strukturę preferencji użytkowników, lecz także uprościć modelowanie, przyspieszyć

trening, umożliwić integrację wielu modalności i poprawić działanie w scenariuszach zimnego startu.

Zgłoszone powyżej uwagi mają przede wszystkim charakter dyskusyjny, uzupełniający i porządkujący. Dotyczą one głównie doprecyzowania zakresu niektórych twierdzeń, szerszego omówienia ograniczeń, dodatkowej analizy porównawczej oraz minimalnej korekty technicznej w treści dokumentu. Nie podważają one zasadniczej wartości rozprawy ani jej wkładu naukowego i aplikacyjnego. W mojej ocenie praca spełnia wymagania stawiane rozprawom doktorskim i zasługuje na ocenę pozytywną.

Konkludując, uważam, że rozprawa doktorska Pana mgr. inż. Michała Daniluka spełnia wymagania stawiane w odpowiednich przepisach rozprawom doktorskim i wobec tego stawiam wniosek o jej dopuszczenie do dalszych, przewidzianych Ustawą, etapów postępowania o nadanie stopnia doktora.

dr hab. inż. Piotr A. Kowalski, prof. AGH